Quality Report Belgian SILC2009

Contents

0. Introduction

1. Indicators

1.1 Overview of common cross-sectional EU indicators based on the cross-sectional component of EU-SILC and equivalised disposable income

2. Accuracy

3. Comparability

4. Coherence

Annex 1: Common cross-sectional EU indicators based on the cross-sectional component of EU-SILC and equivalised disposable income
Annex 2: Questionnaires SILC 2009

0. Introduction

This report contains a description of the accuracy, precision and comparability of the Belgian SILC2009-surveydata. It is structured following the guidelines in the commission regulation (EC) no. 28/2004. This results in three chapters:

1. Indicators
2. Accuracy
3. Comparability

The Questionnaires (in French) can be found in annex to this report (see annex 1).

1. Indicators

Explanation on the calculation of the common cross-sectional EU indicators and equivalised disposable income can be found in document EU-SILC 131-rev/04.

The SAS-applications to calculate the indicators were provided by EUROSTAT. The input data files of the calculation process (household register file, personal register file, household data file and personal data file) are the output files of the Belgium EU-SILC 2009 survey.

An interactive overview of the common cross-sectional EU indicators based on the cross-sectional component of EU-SILC and equivalised disposable income can be found on the Eurostat website: http://epp.eurostat.ec.europa.eu. Additional information for Belgium – with details for the regions for example – can be found on the website of Statistics Belgium: http://statbel.fgov.be/nl/statistieken/cijfers/arbeid_leven/inkomens/armoede.

2 . Accuracy

2.1 Sampling Design

2.1.1 Type of sampling

The Belgian EU-SILC 2009 survey is based on a stratified 2-stage sampling scheme in 2004, followed by rotation since 2005. Rotation allows to replace roughly one fourth of the sample each year. Hence, households (ignoring split-offs) participating in 2009 have been drawn for participation since 2006, 2007, 2008 or 2009.

2.1.2 Stratification

The main stratification criterion is the NUTS2 level. The 11 *sampling strata* are the 10 Belgian provinces (5 in Flanders – coded BE21-BE25 – and 5 in Wallonia – coded BE31 to BE35) and the Brussels Capital Region (BE10).

Further implicit stratification is obtained by sorting PSUs (sub-municipalities) on mean income and sorting SSUs (households) in selected PSUs on age of reference person, as explained in the next section 2.1.3.

2.1.3 Sampling units and 2-stage sampling in 2004

In 2004, when organizing EU-SILC for the first time (ignoring the pilot survey in 2003), 2-stage sampling has been applied in each sampling stratum.

Stage 1 – Primary Sampling Units

The primary sampling units (PSUs) in stage 1 are the municipalities, or parts thereof in the larger ones. In each stratum, the PSUs in the frame are first descendingly sorted by average income; next, a fixed number of times a PSU is drawn according to a systematic PPS (*probability proportional to size*) selection scheme, where size is measured as the number of private households. This systematic sampling method generally causes some PSUs being selected repeatedly (e.g. Schaerbeek, a rather large municipality in stratum BE10, turns out to be drawn 6 times). In total, i.e. in all 11 sampling strata together, 275 PSU draws were made in 2004, once and for all (i.e. for the whole duration of EU-SILC).

Stage 2 – Secondary Sampling Units

The *secondary sampling units* (SSUs) in stage 2 are private households. According to each single PSU draw, a group (generally of fixed size) of households is selected in this stage; notice that a *group* of households corresponds to each PSU *draw*.

In 2004, 40 households have been selected for each PSU draw (i.e. in each group); e.g. in Schaerbeek, 6 times 40 households were drawn. Systematic selection of households has been applied, after sorting the households in selected PSUs by age of reference person. Within each group, the selected households were numbered 1 to 40; households 1-10 constitute the first *rotational group* or *replication*, households 11-20 constitute the second rotational group or replication, and so on. The first replication was meant to participate in 2004 only, the second until 2005, and so on.

The initial household sample in 2004 was *self-weighting*, by the combination of (systematic) PPS sampling of sub-municipalities (PSUs) – size of PSUs being the number of private households – and (systematic) sampling of private households (SSUs), as explained.

2.1.4 Renewal of the sample by rotation, since 2005

Since 2005, a rotation scheme has been applied. Details for each year, from 2005 to 2008, can be found in the corresponding (cross-sectional and/or longitudinal ?) Quality Reports (see …).

The rotation pattern is such that the overlap between samples in any two successive years is roughly 75%, and that the sample is completely renewed after 4 years. Hence four replications or rotational groups in each year, one of which is replaced the year after. Since 2005, each new replication remains in the survey during the next 4 years, and since 2007, each of the four replications is in the survey during four consecutive years.

At the start of 2009, the replication that is in the survey since 2005, is entirely (i.e. irrespective of whether the households are responding or not) dropped. The three replications which entered into the survey in 2006, 2007 and 2008, respectively, are retained (including their split-offs); the households belonging to these three replications will be designated 'old' hereafter.

The supplementary sample, i.e. the new replication that replaces the just dropped replication, is obtained by selecting, for each PSU draw, a fixed number of new households from the corresponding PSU. This selection is done again by systematic sampling, after sorting the households in each PSU on age of reference person. The number of new households for each PSU draw, is determined by considering some (expected) attrition of old households, some (expected) nonresponse for new households, and the required/desired minimum and maximum numbers of responding households, given some precision and budget constraints.

Hence, the (cross-sectional) sample of SILC 2009 consists of

- "old" households: drawn between 2006 and 2008; and
- "new" households: drawn in 2009, staying until 2012.

In the D-file, three variables have been added:

- ✓ DB061 is the identification of the primary sampling units (concatenation of 5 digits for the municipalities and one letter).
- ✓ DB063 is the 'multiplicity', the number of times each PSU was drawn in the sample.
- ✓ DB071 is the order of selection of the new households within each PSU draw or group.

2.1.5 Sample size and allocation criteria

In 2009, 16 new households per group are randomly selected. E.g. in stratum BE10, where 50 groups are managed, $50 \times 16 = 800$ new households are selected. In total 4293 new households are selected in 2009. These households are joined with the 5586 old households that remain from previous years (selected in 2006, 2007 or 2008). Hence 9879 households are invited to participate in 2009. Given some attrition of old

households and nonresponse of new households the number of participating households in 2009 is 6135.

Table 1: sample size and achieved response by NUTS2-units

NUTS2	Name	Old (or strange) hh	New hh	Total hh	Accepted hh (DB135=1)
BE10	Brussels	851	764	1615	832
BE21	Antwerpen	868	676	1544	821
BE22	Limburg	376	251	627	444
BE23	Oost-Vlaanderen	632	499	1131	735
BE24	Vlaams-Brabant	542	403	945	590
BE25	West-Vlaanderen	563	334	897	678
BE31	Brabant Wallon	141	129	270	157
BE32	Hainaut	756	591	1347	897
BE33	Liège	485	392	877	557
BE34	Luxembourg	178	95	273	197
BE35	Namur	194	159	353	227
Total	**Belgium**	5586	4293	9879	6135

2.1.6 Sample distribution over time

2.1.7 Substitutions

No substitution was applied in our survey.

2.1.8 Weightings

Recall that, for the first year of the panel (=SILC 2004 in Belgium), the computation of weights involved three stages (described in 134-04)
   (a)   initial weights
   (b)   weights corrected for nonresponse
   (c)   final (calibrated) weights

For 2009, a distinction has to be made between

"old" households   i.e. households that contain at least one sample person who took part in 2008, and had to be surveyed again in 2009 according to the rotation and tracing rules (excluding the outgoing fourth) (household composition may have changed, whence quotations marks)

"new" households   i.e. households that were drawn for the first time in 2009, among those households not containing any sample person already drawn before

This distinction pertains to initial weights and nonresponse correction
   Since the "old" households are selected indirectly from the 2006, 2007 or 2008 samples, and household composition may have changed, some kind of "weight sharing" must be applied to determine the (2009) initial weights, or rather base

weights. On the other hand, "new" households have their own inclusion probability, whose inverse gives the initial weights;

For the "old" households, (2009) nonresponse=attrition can be linked with (2008) SILC information. For the "new" households, all we can rely upon to explain initial nonresponse is auxiliary information from the Population Register (household size, urban/rural character) and the Financial Statistics (median fiscal income by municipality:)

On the other hand,

Calibration can be done together for "old" and "new" households. With respect to our 2004 model, we decided in 2005 to relax the constraints (basically, calibrating at NUTS1-level instead of NUTS2), in order to decrease the standard deviation of weights.

This introduces the following sections

2.1.8.1 Initial weights for the new households
2.1.8.2 Nonresponse correction for the new households
2.1.8.3 Base weights for the old households
2.1.8.4 Attrition correction for the old households
2.1.8.5 Calibration (all households)

2.1.8.1.   *Initial weights for the new households*

Belgium chose to draw the Primary Sampling Units (= municipalities or parts thereof) "forever", and to rotate the Secondary Sampling Units (=households) within the selected PSU's.

The 2004 PPS two-stage sampling design was self-weighting within each stratum h: x denoting any households in municipality X), we had (in 2004)

P (x drawn) = P(x drawn|X drawn) . P(X drawn) = n_h/N_X . N_X/N_h . $g_h = n_h/N_H$. g_h, where

n_h denotes   the number of households to be drawn in the (selected) PSU (viz. 40)

N_X the number of households in the PSU (in 2004)

N_h the number of households in the stratum (in 2004)

g_h the number of PSU's drawn in the stratum.

(This is an oversimplification, since PSU are drawn with repetition; the selection probability for a PSU should be replaced by the expectation of selection multiplicity, and the term 40 by a multiple depending on the selection multiplicity…but the idea is the same).

In 2009, the picture has become

P (x drawn) = P(x drawn|X drawn) . P(X drawn) = m_h/M_X . N_X/N_h . g_h, where

m_h is   the number of households to be drawn in the (selected) PSU (depending on h)

M_X is   the number of households in the PSU (in 2009)

The factor N_X/M_X indicates the increase-decrease in inclusion probabilities in PSU X (still assuming X has been drawn) between 2009 and 2004.

Now it would seem logical to replace N_X by a smaller number, to account for the households[1] already drawn in 2004, 2005, 2006, 2007or 2008 whence immunized from being drawn again in 2009.

However, the following argument shows that (assuming momentarily that X has been drawn and that the population figures N_X and M_X remain stable) matters are not so easy:

$$P(x \text{ drawn in 2009}) =$$
$$(P(x \text{ drawn in 2009}|x \text{ drawn before}) . P(x \text{ drawn before})) +$$
$$(P(\text{drawn in 2009}|x \text{ not drawn before}) . P(x \text{ not drawn before}),$$

the first term vanishes and the second equals $n_h/(M_X-b) . (N_X-b)/N_h$, where b denotes the number of hh already drawn; since both fraction terms are much larger than b (at least 900 in all selected PSU's), the ratio $(N_X-b)/(M_X-b)$ is (close to 1, and) very close to N_X/M_X. Since the term b is an approximation anyway, we chose to stick to $m_h/M_X . N_X/N_h . g_h$ as inclusion probabilities, and its inverse for initial weights **INIwei=DB080**. Note that, with this concept of DB080, the "new" hh correspond to the total Belgian population (some 4,5 millions private hh); before calibrating, theses weights will be scaled down "to make room" for the old hh; recovering the strange hh means that the sum of the pre-calibration weights will be slightly larger than 4,5 millions (average of g-weights slightly less than 1)

2.1.8.2.Nonresponse correction for the new households

Following Eurostat's suggestion (see Document 065, WEIGHTING II. WEIGHTING FOR THE FIRST YEAR OF EACH SUB-SAMPLE), we replaced the homogeneous response groups (based on household size crossed with urbanity) ratio by a multiple regression model (based on the same dummy variables). By "responding", we mean only those households whose results were accepted (DB135=1). Since 2009 we used logistic regression.

The file was split by NUTS1 and the following variables were used
- Everywhere: Household size, recoded into the four values "one", "two", "three" and "four or more" (so three dummies)
- Out of Brussels: DB100 = urbanity
- In Brussels = BE10: median fiscal income of municipality

The regression produced a new variable "expresp", allowing us to define
NRwei = INIwei/expresp

2.1.8.3 Attrition for the old households

Before "sharing" the 2008 weights, a correction for attrition should be introduced. This year, we elected to perform this correction at the level of individuals, since a 2008 sample person either stays in the panel or leaves it (rotated out, left population, noncontact, refusal or inability to respond, while the structure of a household can change. Note that all household characteristics (e.g. HH020) can be distributed to the members.

---

[1] Perhaps a bit less (households that vanished already subtracted) or a bit more (split households, both components of which stayed in PSU, should be subtracted twice)

We separated the "Children" (for which only basic personal information from the R-file and the distributed H-file is available) from the "Adults" (present in the 2008 P-file as well), i.e. those persons born in 1991 or before.

In the children's model, the following predictors (all, except the last, from the 2008 file – although this does not matter much for group A) were used, grouped by type
  A. individual demographic information: age[2] from RB080, sex = RB090,
  B. housing information: dwelling type = HH010 and tenure = HH020
  C. household type: a limited number of dummies, as there is at least one dependent child;
  D. monetary indicators: we refrained from taking the equivalised income (outliers), but took a transform of it, as well as the dummy "poor or not" and the subjective ability to make ends meet = HS120
  E. sampling and rotation: number of years in panel (from DB075) and urbanisation (=DB100)
  F. one variable (paradata) related to fieldwork in 2009 (computed from HB040 and HB050)

For the adults, the same predictors were used, and moreover
  G. variables from the P-file (related to education level and health);
  H. country of birth (dummy Belgium Yes/No)

---

[2] Let us start with a picture (Z in function of age class, "1" denoting the range 0-4, …, "17" the range "80-84", "18' corresponds to '85 or older", age computed here as 2008-rb080)

The highest 2 scores are depicted in white, the lowest 2 in dark blue. We distinguish two local maxima (one among children 5-9, the other one in the area of "old but not too old") and two local minima (one among "young adults" and one for "very old".

perc resp
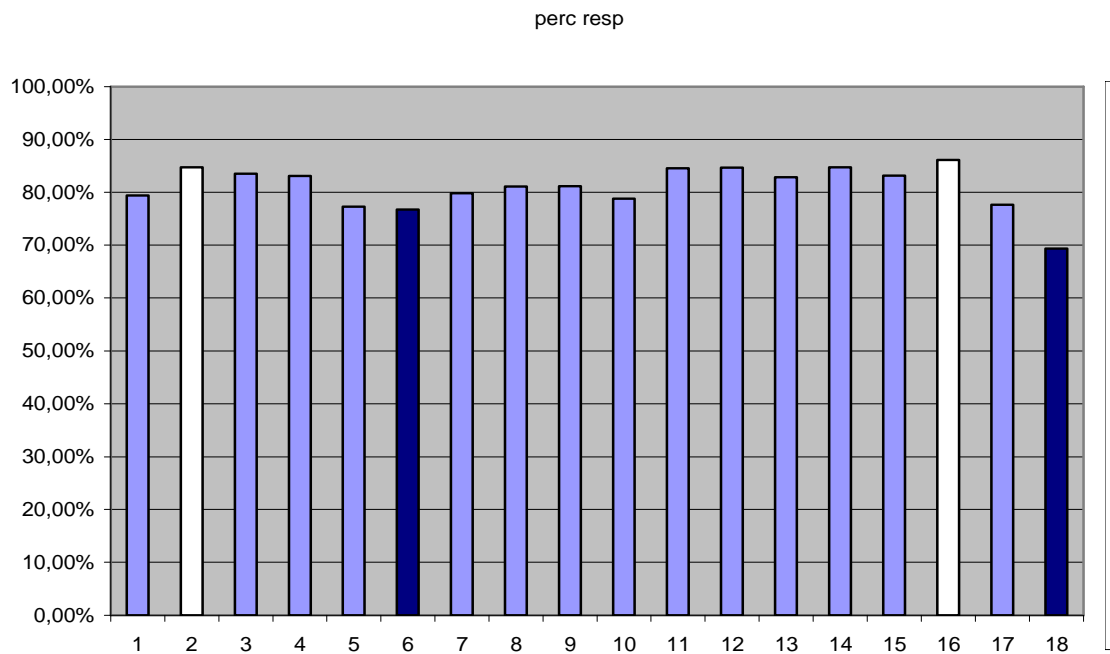
were integrated.

We used logistic regression.

2.1.8.4 Weight sharing

We followed Eurostat's recommendation "EU-SILC weighting procedures: an outline" and shared the calibrated 2008 weights, after correcting for attrition (instead of the initial weights, see Lavallée).

This can be illustrated by an imaginary example, dealing simultaneously with fusions (persons *A&B* in same 2008 hh, *C* in another 2008 hh, so "fusion" in the sense of DB110 occurs), new members (a baby like *E* or already in population like *D*); we focus on the 2009 hh, what happened to those who co-resided with *A* and *B* or with *C* in 2008 (left or split) is irrelevant!
Note that
- RB050 = weight 2008: same for *A* & *B*, vacuous for *D* and *E*
- Newi: in general a bit larger than RB050; *A*'s differs from *B*'s (attrition correction at individual level)
- Somwe = 950+1000+850 involves only *A*, *B* and *C*
- *Weiind:* = ¼ * **somwe** (*A B C D* : four contribute to the denominator)[3]

Person in 2009 hh	A	B	C	D	E
RB110 (2009)	1	1	2	3	4
RB050 (weight 2008)	800	800	600	---	---
Newi = Weight 2007 (after attrition correction)	950	1000	850	---	---
Somwe (sum Newi over 2009 hh)	2800	2800	2800	2800	2800
Weiind	700	700	700	700	700

Weiind will be injected as "initial" weight in the final calibration job.

2.1.8.4 Calibration

We first put the pieces together: weiind is defined as
- (new = started in 2009) :   initial weight, corrected for initial nonresponse, scaled, see 2.1.8.1)
- (old = took part in 2008)   2008 weight, corrected for attrition and weight sharing if necessary, see 2.1.8.4)

---

[3] Do we abide by the Eurostat rules (starting from base weights, it is unclear whether "their" attrition correction precedes or follows weight sharing) ?
There remain some additional categories of persons to be considered:
-Children born to sample women. They receive the weight of the mother (this assumes that the baby belongs to his/her mother's hh)
-Persons moving into sample households from outside the survey population. They receive the average of base weights of existing household members (vacuous here, as RB110 enables us to identify the newborns, but not the immigrants or the –few- persons moving from a collective to a private hh)
-Persons moving into sample households from other non-sample households in the population – these are "co-residents" and are given zero base weight.

- (strange = did not take part in 2008 but before)   initial weight, no correction)

In terms of persons, the weiind statistics were

Type	# ind	Mean of weiind
NEW	4350	602.74
OLD	9670	776.88
BACK	701	371.22
Total	14721	706.11

Recall that 11 *sampling* strata were used (provinces= NUTS2); we use 3 *extrapolation* strata (the 3 NUTS1 regions BRUssels=BE1, VLAanderen=BE2 and WALlonia=BE3)

Calibration model
VLA, WAL:
SIZE4+(AGE8XSEX2)+PROV5   →20 individual[4] + 4 household constraints
BRU:
SIZE4+(AGE8XSEX2)   →16 individual + 4 household constraints

Prov = province where interviewed (differs from DB040 in two cases)

Individual constraints   27=16+11   (age*sex + prov; note that each province belongs to one single region (extrapolation stratum), for the other two regions, the total is set to 0 and the condition is vacuous)

Household constraints   4   (size: "1", "2", "3 or "4 & more",)

Calibration type (after some trials and errors…): truncated

2.1.8.7 Final longitudinal weights

2.1.8.8. Final cross-sectional weights
Statistics

	N	Minimum	Maximum	Mean	Std. Dev.
Final weights	6135	83.68	3898.49	755.45	381.01

2.1.9 Substitutions

No substitution was applied in our survey.

---

[4] Five provinces and 16 age*sex categories, but sum over provinces = sum over age*sex

2.2 Sampling errors

2.2.1 Standard errors and effective sample size

In table 2 we will present an overview of the **standard errors** for the common cross-sectional EU indicators and equivalised disposable income.
An overview of the achieved **sample size** for the 'Laeken indicators' and equivalised disposable income can be found in table 8 of §2.3.3.6.

The design effect for the Median equivalised disposable income = 1.125

2.3 Non-sampling errors

2.3.1 Sampling frame and coverage errors

The sampling frame is the Central Population Register. This Register includes all private households and their current members residing in the territory. Persons living in collective households and in institutions are excluded from the target population. The Central Population Register of 1 February was used.

Updating actions: Central Population Register is updated two times during a month. The changes were communicated to the interviewers.

As there was a period of one month between the drawing of households and the survey itself, over-coverage, under-coverage and misclassification could be happen.

Over-coverage: Persons who died before the survey. Households who moved outside Belgium before the survey. Address is not the principal residence.

Under-coverage: Immigrants who came in Belgium before the survey. Persons who moved from a household to create a new household. Diplomats exempt from an inscription in the national register. Refugees on a waiting list.

Misclassification: Household who moved from a region in Belgium to another region of Belgium.

The size of coverage errors is not available but it was obviously small.

2.3.2 Measurement and processing errors

2.3.2.1. Measurement errors

Measurement errors can occur from different sources, such as the survey instrument, the information system, the interviewer, the mode of collection (CAPI interview). We describe here a few elements by which possible measurement errors can be detected

or which show on the other side the efforts taken to avoid as much as possible measurement errors.

- **Questionnaire construction**
  - ✓ The questionnaire of the SILC2009 survey is the result of several steps:

  - ✓ For building up the questionnaire we took the blue print questionnaire of Eurostat as the basis (documents SILC055, SILC065 and EU-SILC65/02 Addendum II). The order of the questions and the groups (themes of) questions is taken from this blue print. The majority of the questions are almost literally copied (and translated), other questions are changed, however, because experiences in Belgium gave better results posing the questions in another way (The questionnaires were developed in collaboration with the universities that have the experience of the ECHP/PSBH project in Belgium).

  - ✓ After each survey an evaluation of the questionnaire was made (detection of the problematic or difficult to answer questions based on the comments of the interviewers and on a study of the item non-response). When building up the SILC2009 questionnaire we took account of this evaluation.

- **Evaluation of the duration of the interview and the level of difficulty of the questions**

At the end of the interview, the household contact person was asked the following two evaluative questions:

> We would like to thank you for your co-operation. We are at the end of the questionnaire.
> For the evaluation of this questionnaire we would like to ask following questions.
>
> 1. How easy or difficult did you find the answering of the questionnaire in general?
> Very difficult (code 1)
> Difficult (code 2)
> Not difficult but neither easy (code 3)
> Easy (code 4)
> Very easy (code 5)
>
> 2. What do you think of the length of the questionnaire?
> Too long (code 1)
> Neither too long neither too short (code 2)
> Too short (code 3)

In tables 3A and 3B the distribution of the answers on these questions are presented.

Table 3A : Opinion on degree of difficulty of the questionnaire

	N	%
Very difficult	20	0.3
Difficult	240	3.9
Neither difficult/ Nor easy	2305	37.6
Easy	3183	51.9
Very easy	365	5.9
missing	21	0.3
total	6135	100.0

Table 3B : Opinion on the duration of the interview

	N	%
Too long	220	3.6
Neither too long/ Neither too short	5773	94.1
Too short	119	1.9
missing	23	0.4
total	6135	100

For the majority of the participating households (58%), the questions were easy or very easy to interpret. For 94% of the households the interview was neither too long, nor too short. This figure is similar to 2007 and 2008.

As an evaluation after the survey we have sent the households and the interviewers each a different evaluation questionnaire. These questionnaires (the French version) can be found in annex to this Quality Report (see annex 3).

- ***Mismatch in time between household composition and household income (see also §3.1)***

A number of inconsistencies result from a mismatch between the composition of the household at the moment of the interview (between April and November of year x) and the income of the previous year (year x-1).

This mismatch can bias the measurement of poverty status in several ways. For example:

- ✓ Persons who were full-time students in year x-1 (and depending on their parents), but were employed at the time of the interview (and living independently in a one person household for example) will report an income equal to 0 in year x-1 and will be wrongly classified as a poor household.

Other examples can also occur for persons where the household composition changed:

- ✓ For a housewife who was married in year x-1, but divorced and is working at the time of the survey there will also be a mismatch
- ✓ For a household which received family allowances for a student in year x-1, but where the student is no longer part of the household in year x there will also be a mismatch
- ✓ For a household with a person working in year x-1, but retired at the moment of the survey (in year x) a mismatch will also occur. Take notice of the fact that, as the examples show the bias can go in both directions: under and over reporting of income. In each one of the examples, the choice to situate the income reference period in the past is the cause, however.

- ***Error in the routing***

- ***Interview training (Number of training days and information on the intensity and efficiency of interview training)***

Overall we had the impression that the working-experience of the interviewers with EU-SILC starts to pay of. In our opinion the basis data were better for SILC 2009 than

for previous waves. All new interviewers have to follow a two day formation. All trained interviewers followed a formation for an hour and half.

They both had to complete a test-interview before they could download their data. So we can be sure they can completely manage the use of the PC and that they know the questionnaire before they go on the field.

A <u>training group for new interviewers</u> consisted of minimum 5 to maximum 20 interviewers, and according to the size of the training group there were 1 or 2 trainers. Even though the accent was given to the practical side of the training (getting to know the questions and mastering the CAPI-program by imitating interview situations), three manuals were distributed and explained during the training:
- A general manual ('Manuel general aux enquêteurs') containing information about the objectives of the survey, the organisation of the survey, legal and administrative aspects around the survey, fieldwork aspect (how to contact the household, how to introduce oneself, who answers which questions, time delays, …) and the content of the questionnaires.
- A second manual ('Manuel contenu') with all kinds of additional explanations and examples for certain questions/answers.
- A third manual ('Manuel CAPI') about the use of the portable PC for the SILC Computer Assisted Personal Interviews and about the data entry program itself.

The first day of the training there was half a day for learning about and discussing the first two manuals. In the afternoon the trainees received their laptop and got to know the survey and the tool to carry out the interview in practice. One test-interview was simulated collectively. The second day of the training a small part of the time was dedicated to testing to send the data electronically after carrying out the interview. All the rest of the day interviewers practiced several interviews and interview situations with each other on the basis of household profiles that were given. There was also a lot of time for questions and discussions in between these test-interviews.

At the end of the training sessions the instructors had a good image on the degree in which each interviewer ameliorated during the training and on the degree in which they mastered the work. For certain interviewers two days of training was more than enough to master the work, for others it was necessary that they practiced some more at home on specific aspects of carrying out this survey (for example using of the CAPI-program itself, working on the content of the survey, …). They were recommended to do so before carrying out their first real interview. They were often also recommended to start interviewing one-person households.

A <u>training group for trained interviewers</u> consisted maximum 30 interviewers with two trainers. The accent was also given on the content: questions that changed, the module 2009 and questions, which are misunderstood by the interviewers. We made an extra manual for trained interviewers. The trained interviewers obtained four manuals:

- A general manual ('Manuel general aux enquêteurs') containing information about the objectives of the survey, the organisation of the survey, legal and administrative aspects around the survey, fieldwork

aspect (how to contact the household, how to introduce oneself, who answers which questions, time delays, …) and the content of the questionnaires.
- A second manual ('Manuel contenu') with all kinds of additional explanations and examples for certain questions/answers.
- A third manual ('Manuel CAPI') about the use of the portable PC for the SILC Computer Assisted Personal Interviews and about the data entry program itself.
- A fourth manual ('Modifications du questionnaire : module 2009) about the module, changed questions and questions misunderstood by the interviewers.

● *Skills testing before starting the fieldwork*
Interviewers were selected from the interviewer database that Statistics Belgium has centralised for all the survey's that are carried out by the institute. For each interviewer a basic curriculum vitae is present in the database (mentioning for example for which surveys they have experience, their language knowledge, their knowledge of pc, …). A specific unit at Statistics Belgium ('Unité Corps Enquêteurs') is occupied with the selection of the interviewers for each survey; they have good contact with and knowledge of the interviewers. They try to find the best interviewer for each of the geographical areas to cover for SILC. This is not always an easy task because for certain geographical areas several interviewers are candidate, but for other geographical unit there are few or no candidates. Note that interviewers in Belgium most often carry out this work as a second or casual occupation.

● *Skills control during the fieldwork*
During the fieldwork we controlled the work of the interviewers by looking at some of their completed questionnaires. We gave extra attention to all new interviewers and to some trained interviewers that we suspected to be less accurate. Remarks (positive as negative) resulting from these controls were immediately communicated to the interviewer so they could improve their way of working and interviewing.

● *Number of households by interviewer*
Groups of secondary units consisted of about 35 households, depending on the strata. Most of the interviewers had one group of households. Nevertheless several interviewers also had more groups:
interviewers with 1 group: 73
interviewers with 2 groups: 36
interviewers with 3 groups: 21
interviewers with 4 groups: 3
interviewers with 5 groups: 8
interviewers with 6 groups: 1
interviewers with 7 groups: 1

2.3.2.2. Processing errors

Belgium used the CAPI–method to interview the persons. The questionnaire was programmed in Blaise. So processing errors due to data entry (from a written to an electronic format) were reduced to a minimum.
Statistics Belgium programmes several data entry and coding controls in the Blaise program. Below an overview of both data entry and coding controls is presented.

● *Data entry controls*

Table 4: Overview of data entry controls

Question number	Control	Remarks
Contact form		
Column 21, 22, 23 and 24	You can't combine father, mother or being spouse with 'being younger than 12 years".	
Column 8,21 and 22	It's not possible to combine being 'female' and being 'father'. It's not possible to combine being 'male' and being 'mother'.	
Column 21 and 22	Mother and father have to be older than their children (and at least being older than 12 years).	
Column 21, 22, 23, 24	Parents of the spouses or of the partners must be different.	
Column 23, 24	You can't mix 'spouse 'and 'partner'. Must choose one of both for the couple.	
Household questionnaire		
H5 and H7:	It is not possible to combine H5, code 6 with H7 code 2, 3, 4, 5, 6, 7, 8, 9, 10	
H13	Enter a numeric value between 1900 and 2008	
H19	The first of the reimbursement must be between 1954 and 2008 (included). The year of the first purchase must be at the same time or later than the date of buying.	
H27, category g, H45 category g:	Code 1 is only possible if at question H5, code 3,4,5,6 or 7	
H44	Not possible to answer more than 12 months	
H95	Persons have to be between the age of 11 and 23 (included) to obtain a scholarship for secondary school	

H97	Persons have to be between the age of 16 and 99 (included) to obtain a scholarship for higher education	

Individual questionnaire		
Question I6, I7 and I8	You can't combine code 2 of questions I6 and I7 with code 1, 2, 3, 4 and 10 of the question I8.	
Question I6 , I7 and I8	You can't combine code 1 of question I6 or question I7 with code 5, 6, 7, 8, 9 and 11 of the question I8.	
Question I13 and I14:	You can't combine code 1,2,3,4 and 10 question in I13 with code 2 and 3 in question I14	
Question I13 et I16	You can't combine code 1, 2, 3, 4 and 10 of the question I13 with code 1, 2 of the question I16.	
Question I14 and I16	You can't combine code 2 or 3 of the question I14 and code 3 or 4 of the question I16.	
Question I21 and I22	You can't combine code 1,2,3,4 or 10 in question I21 with code 2 or 3 in question I22.	
Question I21 and I29.	You can't combine code 1, 2, 3, 5, 6 of the question I29 with the code 1, 2, 3, 4 or 10 of the question I21.	
Question I29 and I22	You can't combine code 7 of the question I 29 with code 2 or 3 of the question I22.	
Question I37	Age has to be less than current age and not less than 8 year.	
Question I38	Number of years can't be higher than current age minus the age mentioned in question I37.	
Question I 52, I 92.	Can't be higher than 12 months.	
Question I 116	Can't enter a year which is before date of birth.	
Question I25 (I26) (gross income) and question I27 (I28) (net income)	Amounts given in question I25 can't be higher than the amounts given in the question I27.	Ditto for the questions I47 (I48) and i50 (I51), I53 and I54, I55 and I56, I90 and I91, and I93 and I94, I98_A, B, C, D, E, F, G, H and I99 and I102_A, B, C, D, E and I115_ A, B, C, D, E and I116_ A, B
Question I25 and I 26	If the person didn't give an exact amount at the question I25, please go to the question I26.	Ditto for the question I27 and I28; I47 and I48; I50 and I51

Next to these controls, some warnings were implemented in order to ask the interviewer to verify the introduced data in the case of abnormally high or low

amounts. A warning is a simple text box with a message such as 'This amount is very low, are you sure the amount is right?' or 'This amount is very high, are you sure the amount is right?'. The interviewer has then to confirm the value or to change it in case of error.

Household questionnaire	
H16	If lower than 500 or higher than 1000000
H22 (monthly)	If lower than 20 or higher than 2000
H22 (half-yearly)	If lower than 100 or higher than 10000
H22 (yearly)	If lower than 200 or higher than 20000
H23 (monthly)	If lower than 20 or higher than 2000
H23 (half-yearly)	If lower than 100 or higher than 10000
H23 (yearly)	If lower than 200 or higher than 20000
H26	If lower than 25 or higher than 5000
H33	If lower than 50 or higher than 10000
H34, H37, H41	If lower than 100 or higher than 5000
H43, H77, H84	If lower than 25 or higher than 1000
H66	If lower than 100 or higher than 25000
H71B	If lower than 25 or higher than 750
H79, H86	If lower than 25 or higher than 1000
H93	If lower than 100 or higher than 1500

Individual questionnaire	
I25, I27, I47, I50, I90, I91	If lower than 500 or higher than 5500
I53, I54, I55, I56, I86, I93, I94	If lower than 6000 or higher than 66000
I58	If higher than 1200
I98B, I98C, I115B, I115C	If higher than 1350
I99, I102B, I102C	If higher than 5400

Some warnings concern other values than amounts. It's the case for H17 when the value is higher than 30 years ('A period of 30 years is really exceptional, are you sure it is right?') and for H18 when the interest equals 0 or is higher than 15.

It's also the case for H90 for households who say they didn't receive family allowance where children are currently living in the household ('Are you sure you didn't receive any family allowance in 2007 (there is a person of less than 18 year in your household)?').

- *Coding controls*

For the questions relating to occupation (ISCO) and the economic activity of the local unit (NACE) of the main job for respondent, the interviewer could directly insert the corresponding code of the Statistics Belgium. If the interviewer didn't know the corresponding code he could look it up in his computer. If he still hesitated, he could enter a brief description beside the code he entered. These comments were compared with the codes after the fieldwork to correct the data if necessary.

- *Other controls and other problems*
  - ✓ We checked the number of minutes to complete the household and the individual questionnaires (see §2.5). The household questionnaire took about 19 minutes and the individual questionnaires together 24 minutes in means.

2.3.3 Non-response errors

2.3.3.1 Achieved sample size

- *Number of households for which an interview is accepted for the database*

Total: 6135

Rotational group breakdown: group 1: 1823
group 2: 1361
group 3: 1597
group 4: 1354

- *Number of persons of 16 years or older who are members of the households for which the interview is accepted, and who completed a personal interview*

Total: 12236

Rotational group breakdown: group 1: 3474
group 2: 2644
group 3: 3058
group 4: 2591

2.3.3.2 Unit non-response

For the total sample (four rotational groups)

- *Household non-response rates (NRh)*

$$NRh = (1-(Ra * Rh)) * 100$$
where
$$Ra = \frac{\text{Number of addresses successfully contacted}}{\text{Number of valid addresses selected}}$$
$$= \frac{\sum[DB120 = 11]}{\sum[DB120 = \text{all}] - \sum[DB120 = 23]} = \frac{9623}{9762 - 81} = 0.9929$$

$$Rh = \frac{\text{Number of household interviews completed and accepted for the database}}{\text{Number of eligible households at contacted addresses}} =$$

$$= \frac{\sum[DB135 = 1]}{\sum[DB130 = all]} = \frac{6300}{9623} = 0.6378$$

NRh=(1-0.9929*0.6547)*100=36.22%

So, the household non-response rate is 36.2%

- *Individual non-response rates (NRp)*

NRp = (1-(Rp))*100

Where

$$Rp = \frac{\text{Number of personal interview completed}}{\text{Number of eligible individuals}}$$

$$= \frac{11767 - 114}{11767} = 0.9903$$

NRp=(1-0.9903)*100=0.97%

So, the individual non-response rate is 0.97%

- *Overall individual non-response rates (*NRp)*

*NRp=(1-(Ra*Rh*Rp))*100=
(1-(0.9929*0.6547*0.9903))*100=35.62%

So, the overall individual non-response rate is 35.62 %.

For the new households (rotational group 1)

- *Household non-response rates (NRh)*

NRh = (1-(Ra * Rh)) * 100
   where

$$Ra = \frac{\text{Number of addresses successfully contacted}}{\text{Number of valid addresses selected}}$$

$$= \frac{\sum[DB120 = 11]}{\sum[DB120 = all] - \sum[DB120 = 23]} = \frac{4187}{4338 - 106} = 0.9894$$

$$Rh = \frac{\text{Number of household interviews completed and accepted for the database}}{\text{Number of eligible households at contacted addresses}} =$$

$$= \frac{\sum [DB135 = 1]}{\sum [DB130 = all]} = \frac{1823}{4187} = 0.4354$$

NRh=(1-0.9894*0.4354)*100=0.5646%

So, the household non-response rate is 56,5%

- *Individual non-response rates (NRp)*

    NRp = (1-(Rp))*100

    Where

    $$Rp = \frac{\text{Number of personal interview completed}}{\text{Number of eligible individuals}}$$

    $$= \frac{2591 - 38}{2591} = 0.9853$$

    NRp=(1-0.9853)*100=1.47%

    So, the individual non-response rate is 1.47%

- *Overall individual non-response rates (*NRp)*

    *NRp=(1-(Ra*Rh*Rp))*100=
    (1-(0.9894*0.435*0.9853))*100=57.59

    So, the overall individual non-response rate is 57,6 %.

2.3.3.3 <u>Distribution of households by 'record of contact at address' (DB120), by 'household questionnaire result' (DB130) and by 'household interview acceptance' (DB135)</u>

Table 6A: Distribution of households by 'record of contact at address' (DB120), by 'household questionnaire result' (DB130) and by 'household interview acceptance' (DB135)

	Number	Percentage %	Group1		Group2		Group3		Group4	
Total (DB120 =11 to 23)	9879	100,00	4338	100,00	1695	100,00	2150	100,00	1696	100,00
Address contacted (DB120 =11)	9619	97.37	4187	96.52	1679	99.06	2082	96.84	1671	98.53
Address non-contacted (DB120 =21 to 23)	260	2.63	151	3.48	16	0,94	68	3.16	25	1.47
Total address non-contacted	260	100,00	151	100	16	100	68	100	25	100
Address cannot be located (DB120 =21)	53	20.38	32	21.19	6	37.5	12	17.65	3	12,00
Address unable to access (DB120 =22)	16	6.15	13	8.61	0	0	3	4.41	0	0
Address does not exist (DB120 =23)	191	73.46	106	70.2	10	62.5	53	77.94	22	88.00

Table 6B: Distribution of households by 'household questionnaire result' (DB130) and by 'household interview acceptance' (DB135)

	Number	Percentage %	Group1		Group2		Group3		Group4	
Total Household questionnaire completed (DB130 =11)	9619	100,00	4187	100,00	1679	100,00	2082	100,00	1671	100,00
Interview not completed (DB130 =21 to 24)	6137	63.80	1825	43.59	1361	81.06	1597	76.71	1354	81.03
Total interview not completed (DB130 =21 to 24)	3482	36.20	2362	56.41	318	18.94	485	23.29	317	18.97
Refusal to co-operate (DB130 =21)	3482	100,00	2362	100,00	318	100,00	485	100,00	317	100,00
Entire household temporarily away (DB130 =22)	2072	59.51	1509	63.89	149	46.86	244	50.31	170	53.63
Household unable to respond (DB130 =23)	160	4.6	74	3.13	37	11.64	28	5.77	21	6.62
Other reasons	69	1.98	40	1.69	8	2.52	14	2.89	7	2.21
Household questionnaire completed (DB135=1+2)	1181	33.92	739	31.29	124	38.99	199	41.03	119	37.54
	6137	100,00	1825	100,00	1361	100,00	1702	100,00	1576	100,00

Interview accepted for database (DB135=1)	6135	99,92	1823	99,99	1361	100.00	1597	100.00	1354	100.00
Interview rejected (DB135=2)	2	0,08	2	0,01	0	0.00	0	0.00	0	0.00

Longitudinal rate for the 3 groups to follow: 4312/5541=77.82 %

2.3.3.4 Distribution of substituted units

No substitution was applied in our survey.

2.3.3.5 Item non-response

In table 7 an overview of the item non-response for all income variables is presented. The percentage households having received an amount, the percentage of households with missing values and the percentage of households with partial information is calculated.

These percentages are calculated as follows:

- % of households having received an amount : number of households (or persons) who have received something (yes to a filter) / total
- % of households with missing values : number of households (or persons) who said that they have received something but did not give any amount (no partial information) / number of households (or persons) who have received something (yes to a filter)
- % of households with partial information: number of households (or persons – depending on the source of the variable – household file HY or personal file PY) who said that they have received something but gave partial information (amounts were not given for all components) / number of households (or persons) who have received something (yes to a filter)

Table 7: Overview of the non-response for the income variables - % households having received an amount, % of households or persons with missing values and % of households or persons with partial information.

Item non-response	% of households having received an amount	% of households with missing values	% of households with partial information
Total gross household income (HY010)	99,84	11,12	51,71
Total disposable household income (HY020)	99,87	9,14	75,75
Total disposable household income before social transfers except old-age and survivor's benefits (HY022)	97,80	7,17	78,93
Total disposable household income before social transfers including old-age and survivor's benefit (HY023)	95,63	1,5	85,63
Net income components at household level			
Family related allowances (HY050N)	35,75	1,87	1,69

Interests, dividends, etc. (HY090N)	61,61	70,92	0,00
Gross income components at household level			
Income from rental of a property or land (HY040G)	8,74	8,21	0,37
Family related allowances (HY050G)	35,75	1,87	4,15
Social exclusion not elsewhere classified (HY060G)	1,68	0,97	0,00
Housing allowance (HY070G)	0,78	18,75	0,00
Regular inter-household cash transfer received (HY080G)	7,25	1,57	0,00
Interest repayments on mortgage (HY100G)	31,28	13,34	0,00
Income received by people aged < 16 (HY110G)	0,13	0,00	0,00
Regular inter-household cash transfer paid (HY130G)	9,16	1,07	0,00
Tax on income and social contributions (HY140G)	89,80	18,90	55,89
Net income components at personal level	**% of individuals having received an amount**	**% of individuals with missing values**	**% of individuals with partial information**
Employee cash or near cash income (PY010N)	49,15	1,71	1,78
Cash benefits or losses from self-employment (PY050N)	6,11	22,25	0,00
Pension from individual private plans (PY080N)	0,30	8,57	0,00
Unemployment benefits (PY090N)	12,11	7,09	0,00
Old age benefits (PY100N)	19,80	6,48	0,86
Survivor' benefits (PY110N)	0,83	4,08	0,00
Sickness benefits (PY120N)	1,98	6,44	0,00
Disability benefits (PY130N)	3,90	2,40	0,00
Gross income components at personal level			

Employee cash or near cash income (PY010G)	**49,15**	**1,57**	**4,15**
Non cash employee income (PY020G)	**22,42**	**17,21**	**18,20**
Non cash employee income: company car (PY021G)	**4,63**	**27,52**	**1,99**
Cash benefits or losses from self-employment (PY050G)	**6,11**	**56,47**	**0,00**
Pension from individual private plans (PY080G)	**0,3**	**8,57**	**0,00**
Unemployment benefits (PY090G)	**12,11**	**39,65**	**0,00**
Old age benefits (PY100G)	**19,81**	**41,57**	**1,42**
Survivor' benefits (PY110G)	**0,83**	**46,94**	**0,00**
Sickness benefits (PY120G)	**1,98**	**39,91**	**0,00**
Disability benefits (PY130G)	**3,90**	**39,43**	**0,00**
Education-related allowances (PY140G)	**2,20**	**5,02**	**1,16**

2.3.3.6 <u>Total item non-response and number of observations in the sample at unit level of the common cross-sectional European Union indicators based on the cross-sectional component of EU-SILC and for equivalised disposable income</u>
In the table below an overview including interpretation for the non-response is presented.

Table 8: item non-response and number of observations at unit level of the common cross-sectional European Union indicators and for equivalised disposable income.

Indicator	Achieved sample size (number of individuals)	Non-response	remarks
Mean Equivalised disposable income	**14721**	0	-
Risk of poverty threshold: one person household	**1810**	0	
Risk of poverty threshold: household with 2 adults and 2 dependent children	**2728**	0	
Risk of poverty rate by age and gender	**14721**	**0**	
Risk of poverty rate by most frequent activity and gender	**11577**	**190**	

Risk of poverty rate by household type	**14721**	**0**
Risk of poverty rate by household type: Single households	**1810**	**0**
Risk of poverty rate by tenure status	**14721**	**0**
Risk of poverty rate by work intensity of the household	**12987**	**1734**
Dispersion around at risk poverty threshold	**14721**	**0**
Relative median risk-of-poverty gap by age and gender	**2201**	0
Risk-of-poverty rate by age and gender before all transfers (including pensions)	**14721**	**0**
S80/S20 quintile share ratio	**14721**	**0**
Gini coefficient	**14721**	**0**

2.4 Mode of data collection

In tables 9 and 10 the distributions of household members aged 16 and over by 'data status' (RB250) and by 'type of interview'(RB260) are presented.

Table 9: Distribution of household members aged 16 and over by RB250
(Household members RB245=1)

	Total	RB250=11	RB250=14	RB250=21	RB250=23	RB250=31	RB250=32	RB250=33
Total	11767	11653	114	0	0	0	0	0
%	100	99.03	0.97	0	0	0	0	0
Group 1	3474	3436	38	0	0	0	0	0
Group 2	2644	2619	25	0	0	0	0	0
Group 3	3058	3031	27	0	0	0	0	0
Group 4	2591	2567	24	0	0	0	0	0

Table 10: Distribution of household members aged 16 and over by RB260
(Household members RB250=11)

	Total	RB260=2	RB260=5
Total	11652	10151	1501
%	100	87.12	12.88
Group 1	3436	3004	432
Group 2	2619	2265	354
Group 3	3030	2624	406
Group 4	2567	2258	309

2.5 Interview duration

Mean interview duration per household: 46 min (45.77)

2.6 Imputation procedure

2.6.0 Preceding important remark

In contrast to 2004 and 2005 – from 2006 onwards (so also in 2009) the calendar question (i40 in the questionnaire) was presented to every respondent rather the only those who indicated that had been a change in their social-economic position. It enabled us to assess and check much thoroughly the link between the social-economic position and the income variables. Notably for the self-employed this resulted in a substantive number of cases (being identified as being self-employed) who would be otherwise (and who were to some extent in 2004) not identified as being self-employed. These cases mainly concern people in jobs 'somewhere on the bridge' between being self-employed and employee but who nevertheless indicated in the calendar that they were self-employed.

2.6.1 Overall strategy: Emphasis on internal information and integration of outlier detection- , imputation- and control-phases.

Between 2007 and 2008 there was no major changes in our overall strategy.

- Emphasis on internal information.

We can't emphasise enough that to correct and impute our data (for any variable) we relied:
1) **as much as possible on internal information present in the data itself**
2) on formal and legal sources of information and
3) only as final resort turned to statistical procedures (random imputations for ex.)

- An integrated strategy.

As it was the case from SILC 2004 we used again an 'integrated approach' to organise the detection of outliers and the imputations. Crucial to the understanding of our way of working are the concepts of what we call 'vertical' and 'horizontal integration'.

By 'vertical integration' we mean that the phases of outlier detection and imputation were done together for each variable separately (1) rather than that both phases were done separately for all variables together (2). The differences between (1) – the way we did things for SILC 2004 - and (2) the way it was done for SILC 2003 – are subtle but nevertheless more than semantics, especially when combined with horizontal integration.

By horizontal integration we mean that information for each respondent on one variable was checked against information on another variable or another source. Information on the monthly gross income for example was – if both possible and applicable- checked with information on the net income, the yearly income, the current income (if no changes had occurred), the household income, other 'proxi'-variables to income (status etc…) and very important external sources of information like legislation.

The interplay between what we call vertical and horizontal integration leads to a dynamic strategy: variables are checked for outliers and inconsistencies, variables are compared to each other and corrected, (corrected) variables are immediately imputed consistently to the information in other (also corrected) variables – and this several times repeated.

We believe that the emphasis of this strategy on consistency of internal information for respondents throughout the survey and the use of external sources of information (legislation) is a far more successful way of detecting outliers and imputing missing values compared to methods of screening for outliers entirely based on (univariate) distributional features of variables (box-plot methods for example) and imputation methods mainly based on statistical probability models (IVE for example).

Outlier detection: The shift in strategy also implies – of course - a shift in the techniques that are used. As far as the outlier detection concerns there is far less

emphasis on univariate - purely distributional related methods like box-plots but more emphasis on inconsistency checks. For the income variables these checks were done in 2 ways: i. comparison of ratio's between variables and ii. comparison of the relative position of a respondent's answer on one variable to its position on another variable.

i.   Comparison of ratio's between variables:

Comparison of the ratio between two inputs on comparable income variables is a straightforward way to detect outliers. Atypical large or small ratios between gross and net variants of income variables are obviously an indication of 'something being wrong'.

ii.   Comparison of relative positions on income variables:

The central issue in this procedure is the comparison of two income variables by comparison of the normal scores calculated for each case on both variables, after log-transformation. The log-transformation is necessary to normalize the otherwise poisson-distributed income variables.
The inputs of both comparable incomes are considered to be consistent if both normal scores are within predefined boundaries (for example -1,96 and 1,96) and/or the difference between the normal scores is limited (less than 1,96).
There is an indication of bias if the input of one of the incomes for a case is situated within 'normal boundaries' (-1.96 – 1.96) but the other input is not and/or if the difference between the two normal scores differ substantially (>1.96). In fact, the entire procedure consist out of 4 steps:
1. Identification of the variables to be compared.
2. Log-transformations, normality checks, calculation of means and standard deviations.
3. Calculation of normal scores.
4. Consistency control and identification of inconsistencies.

iii.   Other techniques :

There was explicitly more emphasis on the above techniques but this does not imply that the 'conventional' box-plot method was not used at all. In this method input outside the interval below were considered to be outliers:
[First Quartile – 1,5 * (Third Quartile – First Quartile) ; Third Quartile +
1,5 * (Third Quartile – First Quartile)]

Furthermore and as already mentioned, where applicable and usable legal maximums and minimums were also used to some extent.

Finally, we also checked for outliers via controls on a 'case to case' base in which we maximally used information of proxi-variables like professional status and other variables. In this process manifest errors in proxi- and/or other variables associated with the income variables were also removed/corrected (for example 'the number of months').

Imputation: We did no longer make use of IVE. Instead we i. corrected (not imputed – in fact) a greater number of cases and if correction was not desirable or possible, but information on a directly comparable variable was present anyway (see section on internal information above), we ii. resorted to direct imputation, via a regression model.

i.   Corrections.

Corrections were also mainly done on basis of information in other comparable variables. Gross-net ratio of around 40 - 1 Euro = +/- 40Belgian Francs - or 12 - yearly income entered as monthly or vice versa - lead to simple corrections of the gross or the net, for example.

ii.   Regressions.

If correction was not desirable or possible but information on a directly comparable variable was present anyway, we resorted to direct imputation, via a regression model, of the variable for which input was missing. Below we describe how this was done for net –gross imputation, which were the most prevalent instances of that sort. The method was extended, however, to other imputations (imputations of the 2008 income based on the current income, for example).

Missing values on gross income variables (PY010G, PY020G, … and components) were, if collected, imputed on the basis of the corresponding net variables (PY010N, PY020N, … and components). The implementation of this imputation procedure was quasi-similar for almost all (income) variables on which it was applied. The procedure implied 6-steps:

1. Identification of the 'reference cases' (both gross and net collected) and identification of the cases to be imputed (net collected – gross missing).
2. Calculation of the gross/net ratio for the reference cases. Cases with an extreme value on this ratio were excluded from further use in the procedure.
3. Curve estimation of the relation (regression model) between gross and net income. The best fitting model (linear, logarithmic, quadratic, exponential) was being implemented.
4. Implementation of the regression model for the reference cases to identify outliers.
5. Re-implementation of the regression model for the reference cases after removal of the outliers.
6. Actual imputation step: missing (gross) values are imputed on the basis of
   a) net values and
   b) the estimates for the relation between gross and net income assessed in the steps above.

In step 1 the cases of which both gross and net income were collected are identified. We refer to these cases as 'reference cases' (step 1). The relationship between their net and gross income serves as reference for the

imputation of the gross incomes for the cases where only the net was collected (cases to be imputed).

To avoid bias in this imputation model atypical reference cases (both outliers and errors) were identified and removed at several steps in the procedure (step 2 and 4).

In step 2 (reference)cases for whom the ratio between gross and net income exceeded what can be considered typical for the taxation regime applicable to the income concerned, were excluded.
In the case of almost all variables the boundary value of this ratio was set at 2,5. This boundary was arbitrary chosen.
Scrutiny of the excluded cases, however, validates this value's potential to discriminate between incomes which were subjected to real(istic) taxation and outliers or errors.
The latter category seldom counted more than a few percent of the total population in the survey and their gross/net ratio often exceeded the 2,5 considerably.
Further exploration also revealed that the exclusion of these cases from the procedure results in a dramatic increase of the fit of the regression model on which the imputation is based.

In step 4 outliers in the regression model were identified and removed using default regression diagnostics.

The underlying probability model of the net-gross relation was assessed with SPSS' 'curve-estimation' procedure (step 3). It can be hypothesised that in most taxation schemes this relation will not be linear as higher revenues will be subjected to disproportionate higher taxes. The concern therefore is that application of a linear regression model may lead to biased result. Step 3 is an answer to that concern, which turned out to be unfounded, however. In fact, for most variables the linear model fitted the data well. For a few variables the fit of the quadratic model was slightly better, however. Overall, and we underline this, <u>the fit was very good and R-squares very high</u> (always > 0.85)

The estimates of this regression model (step 5) served as direct input for the implementation of the actual imputation (step 6).

iii.   Other techniques.

Although we preferred the techniques above we were in some instances forced to resort to other techniques (due to lack of information – for example).

For some cases we imputed median values calculated after categorising using relevant variables. Most of the median values imputed, were for example, calculated after categorisation for status.

2.6.2 Description on imputation per target variable

In the following table is shown which imputation method we used for each target variable (and also for each component within the Belgian questionnaire). For information on imputation methods please check the document EU-SILC065 where a comprehensive outline is given on the subject. IMD_0 – stands for no imputation. In the majority of the cases of gross variables IMD_2 stands for net- gross imputation.

Table 11A: % Imputationmethod over the total number of observations per (target) variable – gross variables on household level

VARIABLE	IMD_0	IMD_1	IMD_2	IMD_3
HY040G	91,42	0,00	7,65	0,00
HY050G	93,98	6,02	8,58	0,00
HY060G	100,00	0,00	0,00	0,00
HY070G	81,25	0,00	18,75	0,00
HY080G	98,43	0,00	1,57	0,00
HY081G	98,99	0,00	1,01	0,00
HY090G	29,07	0,00	70,93	0,00
HY100G	86,66	-	13,34	-
HY110G	100,00	0,00	0,00	0,00
HY120G	-	-	-	-
HY130G	98,93	0,00	1,07	0,00
HY131G	99,60	0,00	0,4	0,00

Table 11B: % Imputationmethod over the total number of observations per (target) variable – NET variables on household level

VARIABLE	IMD_0	IMD_1	IMD_2	IMD_3
HY040N	-	-	-	-
HY050N	96,44	3,56	0,00	0,00
HY060N	-	-	-	-
HY070N	-	-	-	-
HY080N	-	-	-	-
HY081N	-	-	-	-
HY090N	29,07	0,00	70,93	0,00
HY100N	-	-	-	-
HY110N	-	-	-	-
HY120N	-	-	-	-
HY130N	-	-	-	-
HY131N	-	-	-	-

Table 11C: % Imputationmethod over the total number of observations per (target) variable – gross variables on Personal level

VARIABLE	IMD_0	IMD_1	IMD_2	IMD_3
PY010G	93,57	1,43	0,66	4,34
PY020G	64,59	0,00	12,08	35,41
PY021G	72,48	0,00	27,52	0,00
PY030G	100,00	0,00	0,00	0,00
PY035G	-	-	-	-
PY050G	43,53	0,00	0,14	56,33
PY070G	-	-	-	-
PY080G	91,43	0,00	8,57	0,00
PY090G	99,58	0,28	0,14	0,00
PY100G	57,01	0,47	7,08	35,44
PY110G	53,06	0,00	0,00	46,94
PY120G	60,09	3,00	0,00	36,91
PY130G	60,57	1,96	0,00	37,47

**Table 11D: % Imputationmethod over the total number of observations per (target) variable –
Net variables on Personal level**

VARIABLE	IMD_0	IMD_1	IMD_2	IMD_3
PY010N	96,51	1,56	1,64	0,29
PY020N	83,09	16,91	0,00	0,00
PY021N	72,48	0,00	0,00	27,52
PY030N	-	-	-	-
PY035N	-	-	-	-
PY050N	77,75	5,84	15,44	0,97
PY070N	-	-	-	-
PY080N	91,43	0,00	100,00	0,00
PY090N	99,58	0,28	0,14	0,00
PY100N	92,66	0,47	6,52	0,34
PY110N	95,92	0,00	4,08	0,00
PY120N	93,56	3,86	0,00	2,58
PY130N	97,60	1,96	0,00	0,44

Additional remarks on imputations.

o **Gross/Net imputations.**

For a limited number of monetary variables a limited number of respondents had given only a value for the gross variant of the variable (the opposite – only net is given - occurred much more). For these cases a net value was imputed on basis of the gross using the Belgian rules of taxation. A small number of net- pensions and unemployment benefits were imputed in this way.

All the gross-net imputation for PY100 and PY110 was done following the Belgian taxing rules. We first (1) had to determine the status of the person (isolated or married, with or without dependant children, …), then (2) we applied all the taxing rules including reductions of taxes for e.g. dependant child. (3) Once this model has been applied to gross-net transformation, we could use it for the net to gross (very more useful in fact). To do that, we applied the model on each possible amount as fictive gross amount. As result, we got each possible net amount. We then only had to do the correspondence between net and gross amount.

o **Imputation of 'total housing cost'**

For the calculation of the total housing cost, we examined the current costs for small, average and large usage and used these amounts for both outlier detection and imputation, while taking into account other variables such as the number of household members and the household income. The cost for the water usage for example can be subdivided in subscriber money (fixed) and costs for the actual usage (variable). The

cost for the usage of electricity depends largely whether the heating is electric or not: Singles in an apartment without electric heating consume approximately 600 kWh per year (~ 7 euro), while large consumers with accumulation warmth have an annual usage of approximately 20.000 kWh (~ 240 euro).

2.6.3 Imputation of partial unit non-response

The method chosen for Belgium was imputation of an income for each member of the household who did not answer the questionnaire. Imputation is based on the variable RB210 (basic activity status) of the individual given in the R-file. When the answer is missing or 4 (other inactive person), it is chosen not to impute any income. When available, we preferably used the longitudinal information's from 2008 for imputation. For the other cases the chosen method for imputation was imputation of a sub-category median based on age and sex. Net incomes were computed with a gross to net model, based on the imputed gross incomes.

2.7 Collection variable company Car

Since 2005, we decided to work with **the national rules of the tax authorities**. The benefit for individuals of using a company car for private goals was not directly assessed at the interview but afterwards calculated by applying the applicable taxation rules.
The fiscal benefit of all nature that a person has - due to disposition of a company car for private goals - is calculated by multiplying a fixed amount of kilometres driven for private use by a coefficient. To calculate the latest we need the fiscal cylinder capacity of the car. This fixed amount of kilometres driven for private use is for the tax authorities 5000 km if the distance home-work is less than 25 km, and 7500 if it's more than 25 km.

Since 2005, we asked directly the fiscal cylinder capacity and the distance between work and home. In case of non response of the cylinder capacity, we asked the mark, type and registration year of the car. Than we had to use an imputation method.
Imputation: To calculate the cylinder capacity, we did the following. We assumed that a company car is mostly diesel driven. We looked up for each mark, type and diesel engine what the corresponding cylinder capacity is. If we had several cylinder capacities for the type of the mark, we calculated the weighted mean of the cylinder capacity. If there is not diesel version for a type of car, we did the same logic but than for petrol.

Once we had that we could easily find the corresponding fiscal coefficient. Than we only had to multiply it by the fixed amount of kilometres driven for private use to obtain the fiscal benefit of all nature

Example:

Type of car	Fiscal cylinder capacity	Forfait	Distance home work	Fixed amount	Fiscal benefit of all nature
Smart fortwo	5	0,1898	< 25 km	5000	949 €

Smart fortwo	5	0,1898	> 25 km	7500	1423.5 €

After we calculated the fiscal benefit of all nature for a whole year, we weighted it for respondents who didn't dispose for a whole year of the company car. **The fiscal benefit of all nature is a gross non-cash employee income**.

2.8 Imputed rent

From 2007 onwards a measure for 'imputed rent' needs to add to the data.
In the QR-rapport for the 2007 we extensively reported on the method to calculated imputed rent. In the 2009 operation exactly the same method has been used. Results were very similar.

3. Comparability

3.1 Basic concepts and definitions

The reference population
The reference population is all citizens living officially living at Belgian territory (population de jure). This means that the source of our sample is the central population register. This Register includes all private households and their current members residing in the territory. Persons living in collective households and in institutions are excluded from the target population.
(see also §2.3.1)

The private household definition
The definition of household that Eurostat recommends is used. Household is defined as a person living alone or a group of people who live together in the same dwelling and share expenditures including the joint provision of the essentials of living.

The household membership
The definition of household membership is the same as mentioned in the Eurostat document EU-SILC065/03 about the description of target variables (Chapter 'Units').
All household members of 16 year and older **at the end of the income reference period**, are selected for a personal interview.

The income reference period used
The income reference period is a fixed twelve-month period, namely the previous calendar year. For SILC 2009, the income reference period is the year 2008.

The period for taxes on income and social insurance contributions
This is also fixed twelve-month period, namely the previous calendar year. For SILC 2009, the period is the year 2008.

The lag between the income reference period and current variables
The income reference period is the previous calendar year (year 2008) and the current variables refer to the fieldwork period (April-December2009). Therefore the lag is at minimum 4 months and at maximum 12 months.

The total duration of the data collection of the sample
The fieldwork took mainly place in April, May and June 2009. Some interviews also took place from July to November 2009 but they represent round 10 % of the interviews .

HB050 month of the household interview

		Frequency	Percent	Cumulative Percent
Valid	3.00	0	0	0
	4.00	1717	27,99	27,99
	5.00	2029	33,07	61,06
	6.00	1634	26,63	87,69
	7.00	119	1,94	89,63
	8.00	2	0,03	89,67
	9.00	267	4,35	94,02
	10.00	267	4,35	98,37
	11.00	100	1,63	100
	Total	6300	100.0	

Basic information on activity status during the income reference period

Basic information on activity status during the income reference period was mainly obtained via the calendar question (I40) (in contrast to 2004 where it was obtained by combining the answer for question I8 (PL030) with the answer(s) for question(s) I38 (PL200) and for those with a change I40 (calendar question)). ALSO SEE REMARK 2.6.0.

3.2 Components of income

3.2.1 Differences between the national definitions and standard EU-SILC definitions, and an assessment, if available, of the consequences of the differences mentioned will be reported for the following target variables.

Total household gross income

HY010 = PY010 + PY021G + PY050G + PY090G + PY100G + PY110G + PY120G + PY130G + PY140G + HY040G + HY050G + HY060G + HY070G + HY080G + HY090G + HY110 G.

PY021G only contains the value of company cars and is comparable to the variable PY020G of previous waves of the survey.
PY020G is a new variable from 2008 on which contains in addition to company cars other non-cash income for employees such as luncheon vouchers, goods and services provided free or at reduced price by the employer, …

Total disposable household income

HY020 = HY010 – HY140 – HY130

We didn't take count of HY120G, because regular taxes on wealth do not exist in Belgium.

Total disposable household income, before social transfers other than old age and survivors' benefit

HY022 = HY020 - tnetrans - HY050N - HY060G -HY070G
Tnetrans = PY090N+ PY120N + PY130N + PY140N

Total disposable household income, before social transfers including old age and survivors' benefit
HY023 = HY020 - tnetran2 - HY050N - HY060G -HY070G.
tnetran2 = PY090N+ PY120N + PY130N + PY140G + PY100N + PY110N.

Imputed rent

For more information on how imputed rent was implemented in the Belgian EU-SILC data 2009 – see section

Income from rental of property or land
Asked as Eurostat recommends. Income from rental of a property or land refers to the income received, during the income reference period, from renting a property (for example renting a dwelling –not included in the profit/loss of unincorporated enterprises-, receipts from boarders or lodgers, or rent from land) after deducting costs such as mortgage interest repayments, minor repairs, maintenance, insurance and other charges.

Family/children related allowances
Family / children related allowance includes:
- Income maintenance benefit in the event of childbirth
- Birth grant
- Parental leave benefit
- Family or child allowance.

For the SILC 2009 Belgium asked allowances received from the federal government and also birth grants given by some local authorities and medical organizations.

Social exclusion payments not elsewhere classified
Social benefits in the function 'social exclusion not elsewhere classified' includes for Belgium:
- Income support: periodic payments to people with insufficient resources.
- Other cash benefit: support for destitute and vulnerable persons to help alleviate poverty or assist in difficult situations.
  Belgium only took into account the Benefits paid by the Public Social Welfare Organization (not the benefits paid by private or non profit organizations).

Housing allowances
The housing allowances for Belgium includes:
- Rent benefit
- Benefit to owner–occupiers: a means-tested transfer by a public authority to owner-occupiers to alleviate their current housing costs: in practice help with paying mortgages and/ or interest

It excludes:
- Social housing policy organized through the fiscal system
- All capital transfers (in particular investment grants), for example rehabilitation subsidy and/or a building subsidy.

Regular inter – household cash transfers received

Regular inter-household cash transfers received refer to regular monetary amounts received, during the income reference period, from other households or persons. More precise, we asked for 'alimony and child support' and 'regular cash support'.

Interest, dividends, profit from capital investments in incorporated businesses

Interest, dividends, profits from capital investment in an unincorporated business refer to the amount of interest from assets such as bank accounts, certificates of deposit, bonds, etc, dividends and profits from capital investment in an unincorporated business, in which the person does not work, received during the income reference period less expenses incurred.

Interest paid on mortgage

Interest paid on mortgage refers to the total gross amount, before deducting any tax credit or tax allowance, of mortgage interest on the main residence of the household during income reference period.

Income received by people aged under 16

Income received by people aged under 16 is defined as the gross income received by all household members aged under sixteen during the income reference period. We asked the reference person the annual amount received by all children under 16 together.

Regular taxes on wealth

This variable isn't asked/measured for the SILC2009 in Belgium. The reason is that the regular tax on wealth doesn't exist.

Regular inter-household transfers paid

Regular inter-household transfer paid refers to regular monetary amount paid, during the income reference period, to other households. More precise, we asked for 'alimony and child support' and 'regular cash support'.

Tax on income and social insurance contributions

Tax on income refers to taxes on income, profits and capital gains. They are assessed on the actual or presumed income of individuals, households or tax-unit. They include taxes assessed on holdings of property, land or real estate when these holdings are used as a basis for estimating the income of their owners.

Taxes on income include the sum of the following calculations:
(Gross income from salaries – net income salaries)
+ (Gross income from self-employments – net income from self-employments)
+ (Gross income from pension allowances – net income from pension allowances)
+ (Gross income from disability, illness allowances – net income disability, illness allowances)
+ (Gross income from jobseeker's allowances - net income from jobseeker's allowances)

We have also taken into account of the money that people have received from the taxes or that people have paid to the taxes in 2008 (based on their incomes of the year 2007).

Repayments/receipts for tax adjustments

Repayments/receipts for tax adjustments refer to the money paid to/received from Taxes Authorities related to the income received. This variable is already included in the variable ' tax on income and social contribution' (see above), so Belgium didn't provide this variable.

Cash or near-cash employee income

It includes:
- Wages and salaries paid in cash for time worked or work done in main and any secondary or casual job(s).
- Remuneration for time not worked (e.g. holiday payments)
- Enhanced rates of pay for overtime
- Fees paid to directors of incorporated enterprises
- Commissions, tips and gratuities
- Supplementary payments (e.g. thirteenth month payments, fourteenth month payments)
- Profit sharing and bonuses paid in cash
- Additional payments based on productivity
- Allowances paid for working in remote locations (regarded as part of the conditions of the job)
- Allowances for transport to or from work

Non-cash employee income and company car.

Before SILC 2008 variable PY020 consisted only out of 'company Car'. From 2008 onwards other non-cash income elements needed to be added. 'Company car' itself was recorded in PY021.

In SILC-2009 PY020 consists of the following elements:
* Company car
* free or subsidized meals and luncheon vouchers
* reimbursement or payment of housing-related expenses
* reimbursement or payment of gasoline
* reimbursement or payment of car assurance
* reimbursement or payment of mobile phone costs

Employers' social insurance contribution

The outcome of variable PY030 was the result of the following model:

For blue collar workers:

$$((PY010G*1,08)/100)*50,5$$

And for white collar workers:

$$PY010G/3$$

Both equations were derived from social security rules.

Cash profits or losses from self-employment (including royalties)
It includes:
- Net operating profit or loss accruing to working owners of, or partners in, an unincorporated enterprise, less interest on business loans.
- Royalties earned on writing, inventions, and so on not included in the profit/loss of unincorporated enterprises.
- Rentals from business buildings, vehicles, equipment, etc not included in the profit/loss of unincorporated enterprises, after deduction of related costs such as interest on associated loans, repairs and maintenance and insurance charges.

Value of goods produced for own consumption

This variable is not recorded in the file because the value of goods produced for the own consumption does not constitute a significant component of the income. The importance of the component has been assessed using HBS.

Unemployment benefits
Unemployment benefits include:
- Full unemployment benefits: benefits compensating for loss of earnings
- Partial unemployment benefits
- Early retirement for labour market reasons
- Vocational training allowance
- Mobility and resettlement
- Severance and termination payments
- Redundancy compensation
- Subsistence income for persons entering the labour market

Old-age benefit
Old age benefit includes:
- Old age pensions
- Anticipated old age pensions
- Partial retirement pensions
- Survivor's benefits paid after the standard retirement age
- Disability cash benefits paid after standard retirement age

Survivors' benefits
It includes:
- Survivor's pension
- Death grant
- Other cash benefit

Sickness' benefits
It includes:
Paid sick leave

Disability benefits
It includes:

- disability pension
- early retirement in case of reduced ability to work
- care allowance
- economic integration of the handicapped
- disability benefits to disabled children in their own right
- other cash benefit

Education-related allowances

It includes allowances referring to grants, scholarships and other education help received by students.

However to obtain this variable we asked the information on household level instead of personal level because in Belgium this is paid on household level. Afterwards we attributed this amount to the persons in the individual file.

Capacity to face unexpected financial expenses (HS060) : we take into account the capacity with own means (no borrowing from banks or friends ,…) because these opportunities were the subject of other parts of the question in the Belgian questionnaire .

3.2.2 The source or procedure used for the collection of income variables

The collection of the income variables is by interview. Belgium has no income variables collected from registers for the survey of 2009.

3.2.3 The form in which income variables at component level have been obtained
(e.g. gross, net of taxes on income at source and social contributions, net of tax on income at source, net of social contributions)

Table 12: The form in which income variables at component level have been obtained

Areas	Qr. Block	Target Variable	Unit of measurement	Tax or tax-exempt	If taxable, how the amount is recorded
Employee Income	PY010	Gross Employee Cash or near cash Income in reference period	Individual level	Taxable	Net + gross
	PY020	Gross Non-Cash Employee income (company car, mail tickets)	Individual level	Not taxable (mail tickets are not taxable for the employee and can not be deducted from taxes by the employer) (the company car itself is not taxable but the kilometres that are done for job/work distances and for private distances are taxed: there is always a minimum of 5000 km taxed)	
Self-employment Income	PY050	Gross Cash Income benefits/Losses from self-employment (including profit/loss from unincorporated enterprise, royalties)	Individual level	Taxable For losses, this means a deduction from taxes of this amount can be done on other income posts of that year or on income of the next year)	Net OR gross
Imputed rent[5]	HY030	imputed rent	Household level	-	
Property income	HY090	Interest, dividends, profit from capital investments in unincorporated business	Individual level	Taxable	Net
	HY040	Income from rental of property or land	Household level	Taxable	Gross
	PY080	Regular pension from Private (non-ESSPROS) schemes	Individual level	Taxable	Gross (for the major part of the pensions)
Current transfer received	HY050	Family-related allowances: parental leave benefits	Individual level	Taxable	Net + gross
Social benefits:		Family-related allowances:	Household level	Not taxable	
ESSPROS	HY060	Social assistance	Individual level	Not taxable	

[5] Information on that component is asked because it is important to know if :
- an owner is taxed regarding his tenure status (specific tax on property income)
- a 'rent-free' tenant could be taxed on behalf of the accommodation's owner

	HY070	Housing allowances	Household level	Not taxable	
	PY090	Unemployment Benefits	Individual level	Taxable	Net + gross
	PY100	Old-age benefits	Individual level	Taxable	Net + gross
	PY110	Survivor's Benefits	Individual level	Taxable	Net + gross
	PY120	Sickness Benefits	Individual level	Taxable	Net + gross
	PY130	Invalidity Benefits	Individual level	Taxable	Net + gross
	PY140	Education-related Allowances	Household level	Not taxable	
Regular inter household transfer received	HY080	Regular inter-household cash transfers received	Household level	Not taxable, but taxed if alimentation	Gross
Other income received	HY110	Income received by people aged under 16	Household level	Not taxable	
Interest payments	HY100	Interest repayments on mortgage	Household level	Taxable, this means a deduction from taxes can be done	Gross
Current transfers paid	HY130	Regular inter-household cash transfers paid	Household level	Not taxable or deductible, but taxed if alimentation	Gross

3.2.4 The method used for obtaining income target variables in the required form (i.e. gross values)

See above for information on control, correction, imputation and creation of the gross target variables.

4. Coherence

The results of the Belgian EU-SILC 2009 operation are very similar to those of the 2008 operation.

Annex 1: Common cross-sectional EU indicators based on the cross-sectional component of EU-SILC and equivalised disposable income

SEE the EUROSTAT website.

Annex 2 : Questionnaires SILC 2009

SEE website Belgian statistics.